



Automated acoustic detection of *Vanellus chilensis lampronotus*



Todor D. Ganchev^{a,b,*}, Olaf Jahn^{a,c}, Marinez Isaac Marques^{a,d}, Josiel Maimone de Figueiredo^e, Karl-L. Schuchmann^{a,c,d}

^a National Institute for Science and Technology in Wetlands (INAU), Science without Borders Program, Federal University of Mato Grosso (UFMT), Av. Fernando Corrêa da Costa 2367, Cuiaba-MT, Brazil

^b Department of Electronics, Technical University of Varna, str. Studentska 1, 9010 Varna, Bulgaria

^c Zoological Research Museum A. Koenig (ZFMK), Adenauerallee 160, 53113 Bonn, Germany

^d Institute of Biosciences, UFMT, Av. Fernando Corrêa da Costa 2367, Cuiaba-MT, Brazil

^e Institute of Computation, UFMT, Av. Fernando Corrêa da Costa 2367, Cuiaba-MT, Brazil

ARTICLE INFO

Article history:

Available online 7 April 2015

Keywords:

Computational bioacoustics
Automated acoustic detector
GMM-UBM
Southern Lapwing
Vanellus chilensis lampronotus

ABSTRACT

Traditional human-observer-based biological surveys are expensive. Therefore most biodiversity studies are implemented only periodically, for short periods, and predominantly during daytime and under favorable weather conditions. Automated data acquisition and analysis can overcome these shortcomings and facilitate continuous monitoring. Here we report on the development of an automated acoustic recognizer for Southern Lapwing *Vanellus chilensis lampronotus* vocalizations, a first for this species. The recognizer is a species-specific information retrieval agent, which searches throughout long audio recordings in order to detect and timestamp call events of the target species. The recognizer relies on a log-likelihood ratio estimator, based on a Gaussian Mixture Model–Universal Background Model (GMM–UBM), complemented with purposely-developed temporal post-processing that incorporates domain knowledge about the structure of *V. chilensis* vocalizations. Validation experiments with real-field recordings of complex soundscapes indicate that the recognizer is sensitive enough to register *V. chilensis* call events with sound levels down to –30 dB and recognition accuracy of up to 85.6%, at zero false positive rates. The recognizer is considered a valuable tool for computer-assisted analysis of hourly and daily acoustic activity of *V. chilensis* over extended periods of time, as it offers an indispensable support to long-term monitoring studies and conservation efforts in the Pantanal region.¹

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Birds tend to integrate a wide spectrum of ecological factors and are the best-known indicator group for the conservation and management of natural resources in tropical ecosystems, including wetlands (Stotz, Fitzpatrick, Parker, III, & Moskovits, 1996; Wetlands International, 2010). At the same time, crucial information about the life history and ecology of many tropical wetland birds is lacking, particularly with respect to their response to fluctuating water levels and availability of food resources. The main

difficulty in acquiring such information is that classical survey methods require the long-term involvement of well-trained observers, which is expensive and often logistically prohibitive. Thus, biologists who study the consequences of human activity on certain bird populations would greatly benefit from non-intrusive and cost-effective automated monitoring tools.

Recent technological developments, related to automated information retrieval from audio, have focused on the elaboration of methods in support of biological studies on the presence/absence and activity patterns of rare, threatened, or elusive species (Bardeli et al., 2010; Frommolt & Tauchert, 2014; Henríquez et al., 2014; Huang, Yang, Yang, & Chen, 2009; Potamitis, Ntalampiras, Jahn, & Riede, 2014) and tools in support of the non-intrusive and cost-effective assessment of biodiversity (Aide et al., 2013; Briggs et al., 2012; Digby, Towsey, Bell, & Teal, 2013; Stowell & Plumbly, 2014; Sueur, Pavoine, Hamerlynck, & Duvail, 2008).

Specifically Huang et al. (2009) presented an automated tool for the identification of five anuran species. Their approach relies on

* Corresponding author at: Department of Electronics, Technical University of Varna, str. Studentska 1, 9010 Varna, Bulgaria. Tel.: +359 888 096974.

E-mail addresses: tganchev@ieee.org (T.D. Ganchev), o.jahn@zfmk.de (O. Jahn), marinez@ufmt.br (M.I. Marques), josiel@ic.ufmt.br (J.M. de Figueiredo), klschuchmann@googlemail.com (K.-L. Schuchmann).

¹ INAU Project 3.14 “Monitoring Bioindicators and Migratory Birds in the Pantanal. Applied Acoustics – a Tool for Bio-sustainability Assessment”, INAU Lab. 3 – Biodiversity and Ecological Processes (INAU: www.inau.org.br; 2011–2014) – Program CsF (www.cienciasemfronteiras.gov.br/web/csf).

the algorithm of Härmä (2003) to automatically segment the audio into syllables. Subsequently syllables are then classified. In the experimental validation of their approach, Huang et al. (2009) performed closed-set classification of preselected files, which were assumed to contain sounds of a single species. In that sense, Huang et al. (2009) did not investigate the potential of their method with respect to timestamp determination of individual syllables or series of syllables. Classification accuracy ranged between 82% and 100% for the different species, depending on the classifier used.

Henríquez et al. (2014) studied a semi-automatic system for bat classification. For that purpose they built multiple GMM models for each species and then classified some manually selected audio files, each containing sound emissions of a single bat species. Species-level classification accuracy of 92–99.9% was reported.

Potamitis et al. (2014) developed HMM-based recognizers and investigated the automated detection of American Robin *Turdus migratorius* and Common Kingfisher *Alcedo atthis*. On the song-level detection task they reported 85.1% precision for the former and 84.9% for the latter species. Determination of timestamps of acoustic events and their use for estimating activity patterns of the target species was not discussed in their work.

Frommolt and Tauchert (2014) created a system for semiautomatic recognition of the nocturnal activity of Eurasian Bittern *Botaurus stellaris*. Data acquisition was performed with four omnidirectional microphones. The system functionality included detecting the direction of the sound source, estimating the number of calling animals, identification of the bird species, and determination of the timestamps. The system made use of continuous recordings, acquired only during windless nights. Averaged diagrams of the hourly acoustic activity of the target species were presented for the five-year period 2008–2012. Recognition accuracy of 84.9% was reported at nearly zero false acceptance rates.

A comparison of various machine learning approaches for the classification of bird and amphibian calls was provided by Acevedo, Corrada-Bravo, Corrada-Bravo, Villanueva-Rivera, and Aide (2009) and Henríquez et al. (2014); while Stowell and Plumbley (2010) presented a comprehensive review of methods.

In the present contribution we focus on the development of an automated acoustic recognizer for the Southern Lapwing *Vanellus chilensis*, since we consider it an indicator of certain processes in the Pantanal ecosystem, such as the annual flood-drought cycle. The Pantanal wetlands are increasingly affected by climate change as well as the construction of hydroelectric dams, gold mining activities, erosion, and the use of agrochemicals in Brazil's Cerrado region (Brendle, 2003; Junk & da Cunha, 2005). An understanding of the ongoing environmental changes requires careful monitoring of multiple factors, such as the status of certain animal populations, water levels, contaminant concentrations, and climatic data.

Here we regard the resident *V. chilensis* as a model for other wetland birds and for migratory species, as the lapwing forms noisy flocks and carries out local movements in reaction to the extensive flooding of its breeding habitats during several months of the year. Our principal aim is to study the potential benefits of automated detection of *V. chilensis* acoustic activity, without the presence of human observers in the field. We anticipate that such a technological tool would facilitate the long-term monitoring of *V. chilensis* and provide insights into the biology, ecology, and behavior of this species.

To that end we propose a method for the automated recognition of *V. chilensis* sound events (single-note calls and call series), which also estimates the start and end timestamps of each call event. This method is based on the statistical modeling of the acoustic background and the target species making use of the Gaussian Mixture Model–Universal Background Model (GMM–UBM)

approach. This method was originally developed for the needs of speaker verification (Reynolds, Quatieri, & Dunn, 2000), where decisions are made at the level of complete sentences or entire files. Here we make use of a log-likelihood ratio estimator based on the GMM–UBM approach to search continuous audio recordings for the presence of *V. chilensis* sounds. The output scores of the log-likelihood ratio estimator are post-processed with a purposely-developed temporal smoothing algorithm that incorporates domain knowledge about the structure and duration of *V. chilensis* call events. The latter allows detection of single-note call events and call series as well as the determination of the corresponding timestamps.

In Section 2 we provide a short account of the variety of *V. chilensis* vocalizations and present a comprehensive description of the proposed method and the technological framework on which the acoustic event recognizer is built. There we also describe the datasets and define the experimental protocol used in our work. The performance assessment results of the *V. chilensis* recognizer are discussed in Section 3. The evaluation is carried out on recordings of complex soundscapes collected at one of our monitoring stations in the northern part of the Brazilian Pantanal. Making use of the estimated start and end times of each call event, we present diagrams of the hourly and daily acoustic activity of *V. chilensis* over a complete month of continuous recordings made in 24/7 mode. In Section 4 we discuss the innovative aspects of the present work, clarify the current limitations of the *V. chilensis* recognizer, and point out the future research directions. Finally, in Section 5 we make concluding remarks about the scope and the importance of our work.

2. Materials and methods

In this section we consequently outline the audio datasets used in the present study (Section 2.1), provide a description of the overall architecture and building blocks of the proposed *Vanellus chilensis* recognizer (Section 2.2), and specify the experimental protocol used in the technology evaluation tests (Section 2.3).

2.1. Materials

We first specify the origin of audio recordings used in our study (Sections 2.1.1 and 2.1.2), then describe the data splits employed for the purpose of technology development and evaluation (Section 2.1.3), and finally outline the acoustic characteristics of *V. chilensis* vocalizations (Section 2.1.4).

2.1.1. Study area

The study was carried out in agricultural, natural, and semi-natural habitats of the northern Pantanal region, municipality of Poconé, Mato Grosso, Brazil. The audio recordings analyzed here were collected with Song Meter SM2+ recorders (Wildlife Acoustics²) in the areas of Fazenda Pouso Alegre (–16.50303 S, –56.74533 W; 115–126 m a.s.l.; c. 110 km²; recording period Jul. 2012 through Oct. 2013) and SESC Pantanal Private Natural Heritage Reserve (–16.49879 S, –56.41309 W; 119–131 m a.s.l.; 878.7 km²; recording period Nov. 2013 until to date). These activities were carried out by the Computational Bioacoustics Research Unit³ within the scope of the INAU 3.14 Project “Monitoring Bioindicators and Migratory Birds in the Pantanal” of the *National Institute for Science and Technology in Wetlands*,⁴ aiming at the promotion of Applied Acoustomics as a tool for bio-sustainability assessment (Schuchmann, Marques, Jahn, Ganchev, & de

² Wildlife Acoustics: www.wildlifeacoustics.com

³ CO.BRA: www.ic.ufmt.br/cobra

⁴ INAU: www.inau.org.br

Figueiredo, 2014). The overall goals of the project are to establish and calibrate an automated remote monitoring system for bio-indicators of various sound-producing animal species.

2.1.2. The target species

The bird species Southern Lapwing *Vanellus chilensis* (Molina, 1782) is of strictly Neotropical occurrence. It is distributed from Panama south to Tierra del Fuego, Argentina and is present in all countries of South America (Marchant, Prater, & Hayman, 1986; Wiersma, 1996; Ridgely & Gwynne, 1992). One vagrant was recorded north to Campeche, Mexico (Martin, 1997). Four subspecies of *V. chilensis* are usually recognized (Wiersma, 1996; Santos, 2010). However, differences in morphology and vocalizations suggest that two valid species might be involved, with nominate race *chilensis* and closely related *fretensis* restricted to southern South America, and a separation of *cayennensis*, incorporating *lampronotus*, which are distributed from northern Argentina north to Panama (cf. Internet Bird Collection., 2014). The resident subspecies of Southern Lapwing in our central Brazilian study area is *V. c. lampronotus* (Wagler, 1827).

V. chilensis belongs to the avian Order Charadriiformes, which consists of 17 families and 385 recognized species (Gill & Donsker, 2014). In the northern Pantanal this ground-dwelling lapwing is common in open and semi-open habitats, such as savannah grasslands, pastures, and muddy lake margins but virtually absent from forests and dense scrublands. It is vocal during all months of the year, active during day and night, and gregarious outside the breeding season (Marchant et al., 1986; Maruyama, Cunha, Tizo-Pedroso, & Del-Claro, 2010). Extensive areas of the Pantanal are flooded at the height of the rainy season for several months of the year (Junk, da Silva, da Cunha, & Wantzen, 2011). Our study area is inundated mostly between January and March; however in 2013, when the study at Fazenda Pouso Alegre took place, some recording stations were flooded until the beginning of June. When the water level rises *V. chilensis* is forced to perform short-distance movements as it abandons its inundated breeding areas and gathers in appropriate terra firma habitats. This behavior nicely emulates the seasonal movements of migratory species.

2.1.3. Dataset description⁵

Between July 2012 and October 2014 we implemented continuous monitoring in 24/7 mode and collected about 90 TB of audio recordings (files with duration of 14, 15, and 30 minutes, sampling rate 48 kHz, resolution 16 bits), including one complete annual cycle each at Fazenda Pouso Alegre and SESC Pantanal Private Natural Heritage Reserve. From these recordings we extracted several excerpts which served for technology development and evaluation as follows:

2.1.3.1. Acoustic background model training dataset. For the creation of a balanced acoustic background model we combined two datasets of audio recordings, representative of the Pantanal soundscapes. The first consists of fifty-four 14-minute recordings of the Fazenda Pouso Alegre with a total duration of over 12 hours. The original recordings were manually edited to remove over 150 (mostly weak) call series of the target species (*V. chilensis*). The second dataset consists of thirty-two 30-minute recordings made at SESC Pantanal, corresponding to approximately 15 h of continuous recordings from a sound recording station located inside forest. Since the latter habitat is avoided by the target species, the

recordings from that station were regarded as *V. chilensis* free, although occasionally some vocalizing lapwings may have flown over the site. In this way, approximately 27 h of audio recordings from the Pantanal were available for training the acoustic background model.

2.1.3.2. *Vanellus chilensis* training dataset. From about 4500 manually selected snippets of animal sounds we built a reference library of 258 recordings with *V. chilensis* vocalizations. These recordings are excerpts of the original 14, 15 or 30 minute audio recordings and were neither edited nor filtered. Most recordings contained competing sounds of other sources, such as vocalizations of non-target species, anthropogenic noise, and wind. For the training dataset we selected representative recordings from the reference library and subsequently removed non-target signals by manual editing and selective filtering. The final training dataset consisted of 93 recordings, containing ~45 min of target sounds. Among these are 38 recordings virtually clean of competing interference (~18 min), 52 recordings that still contain some other competitive low-volume background sounds overlapping with the target signals (~24 min), and 3 recordings representing choruses of lapwing flocks (~2 min). We aimed to include examples of a wide spectrum of *V. chilensis* call types but for some rarer call variants we found few examples in the reference library.

2.1.3.3. *Vanellus chilensis* validation dataset. The original audio recordings are often very noisy due to wind, rain, competing sound signals of other species, and interferences from other sources. Furthermore, some calls of the target species may be so faint that they are barely identifiable, even for a bird sound expert. To address the challenges related to manual tagging of real-field recordings we took the following approach:

- (a) We compiled a few dozen soundscapes that were known to contain at least one good call series of *V. chilensis*. By selecting recordings that were not previously processed during the manual snippet selection we made sure that there was no overlap between the audio signals used for recognizer training and evaluation.
- (b) We randomly chose a subset of these audio files in order to annotate the timestamps of *V. chilensis* call events (single-note call or a sequence of calls).
- (c) Only for the purpose of tagging we made use of a graphic equalizer to reduce competing noises such as wind and insects by -48 dB in the frequency ranges [0, 630] Hz and [12.5, 24] kHz. In stronger target signals the lowest and some of the upper harmonics were removed involuntarily during this procedure. However the *V. chilensis* recognizer directly processes the original audio as described in Sections 2.2.1 and 2.2.2.
- (d) The filtered recordings were screened by a bird-sound expert using a headset with integrated manual volume control but without additional software amplification, and simultaneously through visual inspection of the spectrograms in Adobe Audition.
- (e) The bird-sound expert tagged the start and end times of each acoustic event; however, the number of calls per vocalization event was not counted.
- (f) The following rules were applied to separate call events:
 - (1) A pause of at least one second between target signals was used to separate call events of similar sound pressure levels.
 - (2) Abrupt changes in sound pressure levels were used to separate different call series even if there was no pause between the signals;

⁵ According to the principles of reproducible research, we will provide all datasets described here through the website of the Computational Bioacoustics Research Unit (CO.BRA; www.ic.ufmt.br/cobra) of the Federal University of Mato Grosso (UFMT).

- (3) However, call sequences that varied greatly in sound pressure levels over time were not divided into different call series when the calling birds seemed to continuously move (fly) around the recording stations.
- (g) In addition, we noted the compound dB-values of the loudest call of each vocalization event; that is, the volume of the target signal plus the volume of the remaining background noise. In most cases this approach led to a considerable overestimation of the dB-values for weaker target signals below about -35 dB. These levels depend on the proximity and orientation of the sound source to the microphone.
- (h) The manual annotation of the timestamps was repeated twice. Depending on the complexity of the soundscape, the time effort per 14-minute sound file was between about 4 and 16 h.
- (i) Some audio signals could not be identified with certainty as *target* or *non-target*, either because they were too faint or because background noise levels were too high. Such call series were time-stamped and excluded from analysis; for dataset (I) these were $N = 70$ and for dataset (II) $N = 137$.

The aforementioned procedure led to two datasets consisting of 14-min original soundscapes containing vocalizations of the target species:

- (I) four recordings with a total of 80 *V. chilensis* call events for the purpose of technology optimization and adjustment of the parameters of the recognizer, hereafter referred to as dataset VL01, and
- (II) ten recordings with a total of 337 *V. chilensis* call events to evaluate the performance of the recognizer, hereafter referred to as dataset VL02.

2.1.4. Acoustic characteristics of *V. chilensis lampronotus* vocalizations

V. chilensis lampronotus vocalizes from the ground and in flight, particularly during courtship and territory displays as well as in response to the presence of potential predators. The most often heard vocalization is a strident “*keh-keh-keh-keh-keh...*” that is consistently repeated by alarmed birds (Gwynne, Ridgely, Tudor, & Argel, 2010, p. 99). On occasions the lapwings give emphatic virtually double-note “*keh-a, keh-a, keh-a, keh-a, keh-a...*” vocalizations or similar variants of the call (Fig. 1, Table 1). Vocalizations of perched birds are often rather short, sometimes consisting of only one or few notes. During the non-breeding season, flocks of a few to several dozen birds congregate and often vocalize in choruses.

The spectral structure of single call notes is quite complex, consisting of multiple harmonics with most of the energy concentrated in the frequency range [1.1, 10] kHz (Fig. 1). The specific peak frequencies of individual call notes seem to depend on the agitation level of the birds (Table 1). In addition, some weak spectral details are available for frequencies below 1 kHz and above 10 kHz, so harmonics may spread over the entire audible frequency range above 550 Hz in recordings made close to the birds (Fig. 1g and h). However, since the energy of these details is weak they are usually buried in the ambient noise, and thus do not offer reliable cues for the acoustic detection of *V. chilensis* in real-field recordings.

2.2. The *Vanellus chilensis* recognizer

The overall architecture of the *V. chilensis* acoustic recognizer is presented in Fig. 2. Two main processing stages are defined: audio parameterization (details in Section 2.2.1) and pattern recognition (details in Sections 2.2.2 and 2.2.3). Firstly, the audio signal is pre-processed and then audio descriptors are computed and standardized to facilitate the pattern recognition process. Secondly, a

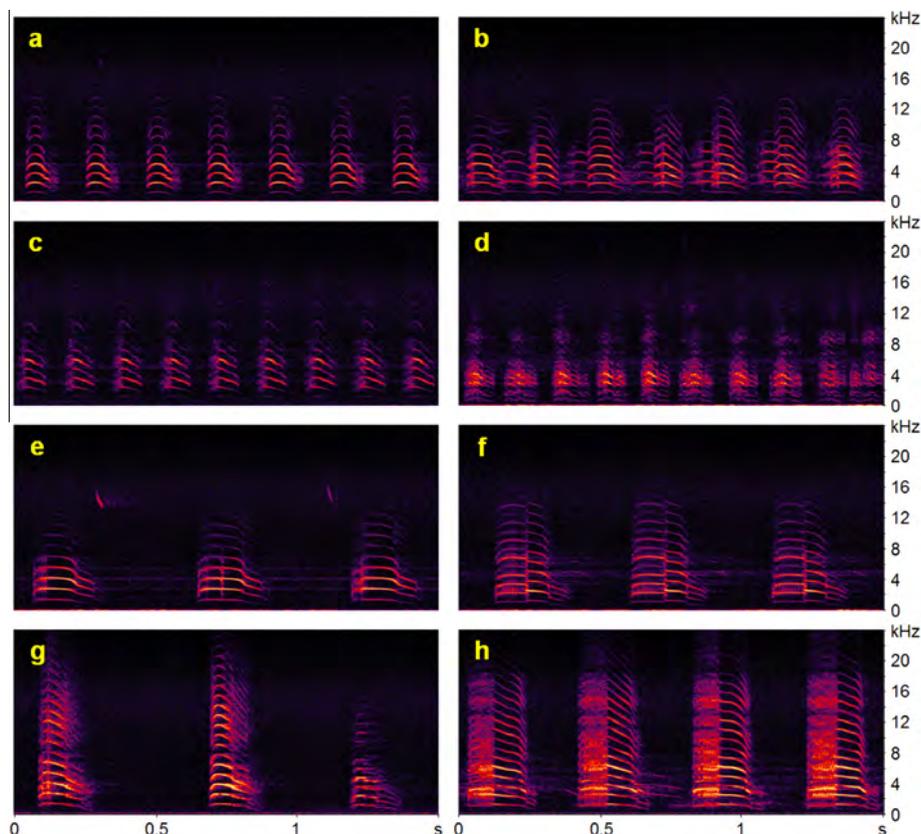


Fig. 1. Examples of *Vanellus chilensis lampronotus* vocalizations from the northern Pantanal, Brazil. Shown are 1.5-second segments of unedited and unfiltered reference recordings sampled at 48 kHz (cf. Table 1 and Appendix A for further details).

Table 1
Characterization of *Vanellus chilensis lampronotus* vocalizations from the northern Pantanal, Brazil (cf. Fig. 1 and Appendix A). The column “*N* calls” refers to the number of calls considered for analysis. The association between vocalizations and behavioral contexts was not previously studied by our project.

Spectrogram	Signals/s	Call duration (s)	Peak frequency (Hz)	N calls	Sound description	Observations
a	4.6	0.123–0.129	2400	50	Fast <i>Accipiter</i> -like series of <i>keh-keh-keh-keh-keh</i> calls	Single perched bird; mate vocalizes shortly thereafter; recorded during night at 02:15 h
b	4.3	0.098–0.135	3400	11	Fast series of <i>keh-keh-keh-keh-keh</i> calls	Chorus of at least four agitated birds in flight; at dawn (05:45 h)
c	5.9	0.107–0.116	6000	17	Fast and high-pitched series of <i>kre-kre-kre-kre-kre</i> calls	Single agitated and perched bird in response to a short call series of a conspecific individual
d	6.3	0.112–0.136	3000/4000	9	Fast series of rail-like <i>kreh-kreh-kreh-kreh-kreh</i> calls	Two agitated birds at dusk
e	1.9	0.240–0.260	2800	5	Slow series of <i>keeh, keeh, keeh, keeh, keeh</i> calls	Single bird; second bird vocalizes shortly thereafter; recorded during night at 01:30 h
f	2.0	0.241–0.273	2300	23	Slow series of emphatic virtually double-note <i>keh-a, keh-a, keh-a, keh-a</i> calls	Single perched bird; mate vocalizes shortly before and after; recorded during night at 02:30 h
g	2.0	0.192–0.235	3800	7	Emphatic <i>keeh-keeh-keeh-keeh-kwo</i> calls	Flight calls of a single bird; a second bird vocalizes shortly before and after
h	2.5	0.218–0.264	2400	10	Emphatic <i>kweh-kweh-kweh-kweh-kweh</i> calls	Flight calls of a single bird; the mate and other conspecific birds vocalize shortly before and after

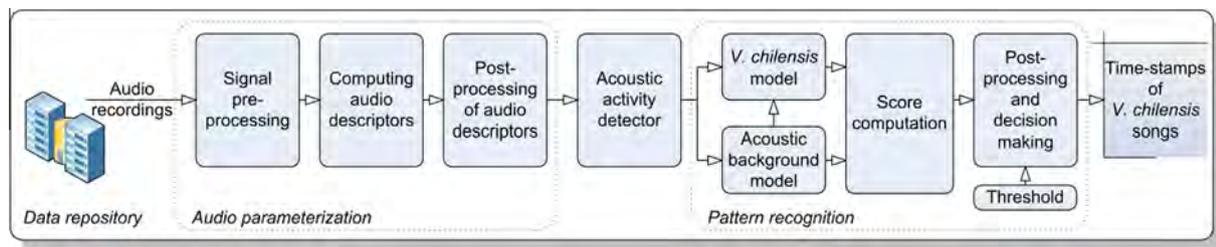


Fig. 2. Overall block diagram of the *Vanellus chilensis* recognizer.

detector is used for selecting those parts of the audio recording where acoustic activity occurs and for labeling the signals for further processing. The pattern recognition stage evaluates the degree of proximity between the audio feature vectors and the *V. chilensis* model and the acoustic background model. The resultant scores are post-processed and subsequently serve as the basis for decision making about the presence or absence of *V. chilensis* vocalizations in the specific portion of audio. The final recognizer output consists of timestamps that specify the start and end times of each *V. chilensis* call event.

2.2.1. Audio parameterization

The audio parameterization converts the acoustic signal into a series of low-dimensional feature vectors, each representing a short segment of the acoustic signal. The block diagram, shown in Fig. 3, summarizes the audio parameterization steps.

The first step, denoted as *pre-filtering* of the input signal $s(n)$, consist of mean value removal, band-pass filtering, and down-sampling of the signal to 24 kHz in order to reduce the computational demands during the subsequent processing steps. Band-pass filtering with Butterworth filter of order 10 is seen as a simple but adequate noise reduction step, which aims at improving the signal-to-noise ratio (SNR) before parameterization. It preserves the frequency range of 800 Hz to 11 kHz, where most of the energy of *V. chilensis* vocalizations is located and reduces the influence of additive low-frequency interferences (for instance due to wind blowing in the microphone or to mechanical vibrations provoked by wind). Except for the attenuation of the dc-offset and low-frequency interferences, the band-pass filtering also reduces the contribution of the high-frequency components that might occur in clipped signals.

The second processing step, *signal segmentation*, is based on the assumption that the spectral characteristics of the audio signal do not change significantly within short periods of time. In the process of segmentation, we apply a sliding window function with 20 ms in length to the band-limited audio signal. At a sampling frequency of 24 kHz this leads to the creation of audio segments with length $N = 480$ samples. Subsequent segments are overlapped by 75%, i.e., the time step of audio parameterization is 5 ms. The pre-processed signal is subject to parameterization as follows:

Each audio segment, consisting of N samples of the pre-processed signal, is subject to the short-time discrete Fourier transform:

$$S_l(k) = \sum_{i=0}^{N-1} \hat{s}(i+l) \cdot W(i) \cdot \exp\left(-\frac{j2\pi ik}{N}\right), \quad 0 \leq i, k \leq N-1, \quad (1)$$

where i is the index of the time domain samples, k is the index of the Fourier coefficients, and l denotes the relative displacement of the current segment. Since the following steps apply to every segment, in the rest of this article we omit the subscript l but it remains implied. The Hamming window

$$W(m) = 0.54 - 0.46 \cos\left(\frac{2\pi m}{N}\right), \quad m = 0, 1, \dots, N-1, \quad (2)$$

is applied to reduce the spectral distortions caused by an abrupt change of signal amplitude at the boundary points of the audio segment.

We cover the frequency range [1, 11] kHz with the filter bank $H_i(k)$, consisting of $B = 100$ equal-bandwidth and equal-height

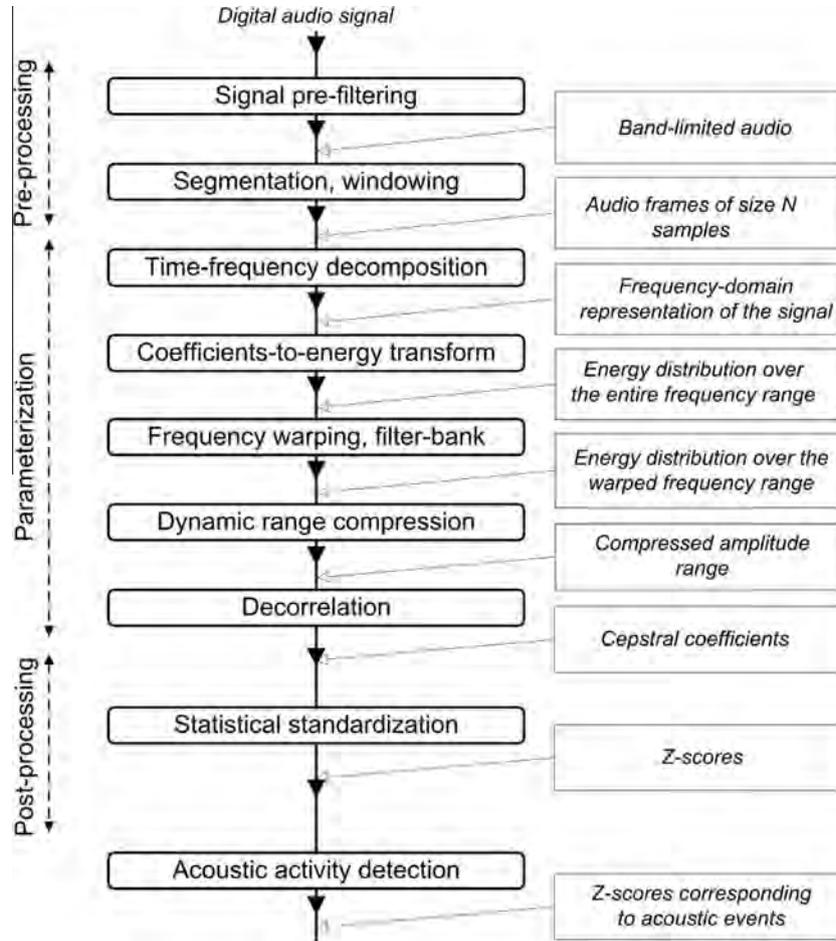


Fig. 3. Block diagram of the audio parameterization process.

filters on the logarithmically compressed power spectrum. Each filter is defined as:

$$H_i(k) = \begin{cases} 0 & \text{for } k < f_{b_{i-1}} \\ \frac{(k-f_{b_{i-1}})}{(f_{b_i}-f_{b_{i-1}})} & \text{for } f_{b_{i-1}} \leq k \leq f_{b_i} \\ \frac{(f_{b_{i+1}}-k)}{(f_{b_{i+1}}-f_{b_i})} & \text{for } f_{b_i} \leq k \leq f_{b_{i+1}} \\ 0 & \text{for } k > f_{b_{i+1}} \end{cases}, \quad (3)$$

where $i = 1, 2, \dots, B$ stands for the i th filter, f_{b_i} are the boundary points of the filters, and $k = 1, 2, \dots, N$ corresponds to the k th coefficient of the N -point discrete Fourier transform (DFT). The boundary points f_{b_i} are expressed in terms of position, which depends on the sampling frequency F_s and the number of points N in the DFT. The center of the first filter is at 1 kHz and the centers of the subsequent filters are linearly displaced at 100 Hz and serve as boundary points for the neighboring filters. We have chosen linear spacing (equal frequency resolution) because *V. chilensis* produces wide-band sounds which are spread over the entire frequency range [1, 11] kHz.

Next, we perform decorrelation of the log-energy filter-bank outputs S_i , where

$$S_i = \log_{10} \left(\sum_{k=0}^{N-1} |S(k)|^2 \cdot H_i(k) \right), \quad (4)$$

via the discrete cosine transform (DCT):

$$x_j = \sum_{i=1}^B S_i \cos \left(j(i-0.5) \frac{\pi}{B} \right), \quad j = 0, 1, \dots, D. \quad (5)$$

In (5) j is the index of the linear frequency cepstral coefficients (LFCC), and D is the total number of LFCC that are computed, where $D \leq B$. In the present work $D = 35$, as these represent over 90% of the signal energy.

Since the above analysis applies to every audio segment, the output of the signal parameterization stage is a sequence of LFCC vectors. We use \mathbf{X}_{all} to represent a sequence of LFCC vectors extracted from the original audio signal. In the general case, \mathbf{X}_{all} is a sequence of T D -dimensional LFCC vectors, $\mathbf{X}_{all} = \{\mathbf{x}_t\}_{t=1}^T$, where $\mathbf{x}_t = \{x_j\}_t$ is the vector at time instance t , and T is the total number of vectors.

Subsequently, *post-processing* of the LFCC is carried out in order to reduce the negative consequences of time-varying acoustic conditions and to exclude from further processing the non-promising regions of the audio. For this purpose we first perform statistical standardization of the LFCC, i.e., \mathbf{X}_{all} is normalized to zero mean value and unit standard deviation. This results in obtaining the z-score \bar{x}_j of each individual parameter x_j of the LFCC vector:

$$\bar{x}_j = \frac{x_j - v_j}{\xi_j}, \quad j = 0, 1, \dots, D. \quad (6)$$

The z-scores, computed from the LFCC, quantify the distance from the mean in terms of the standard deviation. The mean value v_j and the standard deviation ξ_j are estimated on a *per file* basis for each of the D dimensions.

Furthermore we make use of an energy-based acoustic activity detector to prune the sequence of audio feature vectors \mathbf{X}_{all} and exclude from further processing the audio feature vectors $\bar{\mathbf{x}}_t = \{\bar{x}_j\}_t$ corresponding to low-energy frames. The former step helps to remove silence from the training data and to speed up computations in the *V. chilensis* recognizer. Our acoustic activity detector is based on modeling the log-energy distribution on a *per file* basis with a three-component Gaussian Mixture Model (GMM):

$$p(\bar{x}_{0t}; w, \mu, \sigma) = \sum_{i=1}^3 w_i \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_i^2} (\bar{x}_{0t} - \mu_i)' (\bar{x}_{0t} - \mu_i) \right\} \quad (7)$$

where w_i are the mixture weights, μ_i are the means, and σ_i^2 are the variances of the individual components. These parameters are estimated from the sequence of z-scores \bar{x}_{0t} , computed after applying (6) for the 0th LFCC x_0 (5). From (5) we can see that x_0 is proportional to the full-band log energy computed for each audio frame.

The maximum likelihood model parameters of the GMM are estimated using the iterative expectation–maximization (EM) algorithm.

In order to identify the silent audio portions, we interpret the individual components of the three-component GMM as follows:

- The mixture component with the lowest mean value μ_1 corresponds to portions of audio with the lowest energy, and we therefore label these as *silence/background noise*, which is excluded from further processing.
- The mixture component with the highest mean value μ_3 corresponds to portions of audio with significant acoustic activity and therefore we label these as *acoustic events* kept for further processing.
- Finally, the mixture component with the middle mean value μ_2 corresponds to weak signals originating from low-energy sources or sound sources located at a significant distance from the microphone. We keep only the upper P percents of the middle component.

As a result, the acoustic activity detector acts as a gateway which labels for further processing sequences of audio feature vectors $\mathbf{X}_e \subseteq \{\bar{\mathbf{x}}_t\}_p$, each corresponding to an acoustic event with energy in the range specified by the upper P percent of the middle Gaussian component (with mean μ_2) and the third one with mean μ_3 . In the present work we consider $P = 15\%$ for the training dataset and $P = 75\%$ for the tests datasets as we aim at recognizing *V. chilensis* vocalizations at audio levels down to -30 dB. In the following we process each sequence \mathbf{X}_e independently from the others. For simplicity of notations we drop the subscript e but it remains implied.

2.2.2. GMM–UBM-based bird species detection

The *V. chilensis* recognizer follows the GMM-UBM likelihood test approach, introduced by Reynolds et al. (2000) on the speaker verification task. In brief, given a sequence of audio feature vectors \mathbf{X} , computed for a sequence of audio frames and the hypothesized bird species S , the task of automated bird species detection consists in determining whether \mathbf{X} originates from the target species S . From this perspective, the task consists in testing the following two hypotheses:

- H_0 : \mathbf{X} originates from the hypothesized bird species S ,
 H_1 : \mathbf{X} does NOT originate from the hypothesized bird species S .

Given a sequence of audio feature vectors \mathbf{X} , and given a mechanism for the reliable estimation of likelihoods of the hypotheses

H_0 and H_1 , we can apply the optimum test (8) and decide between these two hypotheses:

Likelihood ratio test	Comparison with threshold and decision (8)
$\frac{p(\mathbf{X} H_0)}{p(\mathbf{X} H_1)}$	$\geq \theta$ accept H_0 , (Decision: \mathbf{X} originates from the bird species S), $< \theta$ reject H_0 , (Decision: \mathbf{X} does NOT originate from the bird species S).

Here $p(\mathbf{X}|H_0)$ and $p(\mathbf{X}|H_1)$ are the probability density functions for the hypotheses H_0 and H_1 , which are evaluated for the sequence of audio feature vectors \mathbf{X} , computed for the observed audio segment. The decision threshold θ for accepting or rejecting H_0 needs to be estimated on the basis of a representative development dataset.

In practice, H_0 is represented with the mathematical model λ_{hyp} , which characterizes the hypothesized species S in the audio feature space. Likewise $\lambda_{\overline{hyp}}$ represents the alternative hypothesis H_1 . Thus the likelihood ratio (8) can be rewritten as

$$\frac{p(\mathbf{X}|\lambda_{hyp})}{p(\mathbf{X}|\lambda_{\overline{hyp}})} \quad (9)$$

Furthermore, the logarithm of (9) gives representation of the log-likelihood ratio as:

$$\Lambda(\mathbf{X}) = \log [p(\mathbf{X}|\lambda_{hyp})] - \log [p(\mathbf{X}|\lambda_{\overline{hyp}})] \quad (10)$$

The model λ_{hyp} for H_0 is well defined and can be estimated using the training dataset for the target bird species S . However the model $\lambda_{\overline{hyp}}$ for H_1 is not well specified as it has to represent all possible alternatives to the hypothesized target species.

Given a collection of audio recordings from a large number of species that are representative of the community of sound-emitting species observed in the habitat, a single model $\lambda_{UBM} \sim \lambda_{\overline{hyp}}$ is build to represent the alternative hypothesis. It is also possible to use multiple background models tailored to specific sets of species or habitats, but the use of a single background model has advantages in terms of computational efficiency.

An important step in the implementation of the likelihood ratio test is the selection of the actual likelihood function $p(\bar{\mathbf{x}}_t|\lambda)$. The choice of this function depends on the audio features being used, as well as on the specifics of the application. The mixture density used for the likelihood function is defined as follows:

$$p(\bar{\mathbf{x}}_t|\lambda) = \sum_{i=1}^M w_i p_i(\bar{\mathbf{x}}_t) \quad (11)$$

The mixture density is a weighted linear combination of M unimodal Gaussian densities $p_i(\bar{\mathbf{x}}_t)$, each parameterized by a $D \times 1$ mean vector $\boldsymbol{\mu}_i$ and a $D \times D$ covariance matrix $\boldsymbol{\Sigma}_i$. The mixture weights w_i satisfy the constraint $\sum_{i=1}^M w_i = 1$. Here $p_i(\bar{\mathbf{x}}_t)$ are defined as:

$$p_i(\bar{\mathbf{x}}_t) = \frac{1}{(2\pi)^{D/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\bar{\mathbf{x}}_t - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\bar{\mathbf{x}}_t - \boldsymbol{\mu}_i) \right\} \quad (12)$$

While the general form of the model supports full covariance matrices, typically only diagonal covariance matrices are used in order to reduce the number of adjustable parameters. Collectively, the parameters of the density model are denoted as $\lambda = \{w_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$, $i = 1, 2, \dots, M$.

Given a training dataset of audio feature vectors, maximum likelihood model parameters can be estimated using the iterative expectation-maximization (EM) algorithm. The EM algorithm iteratively refines the GMM parameters to monotonically increase the likelihood of the estimated model for the observed feature vectors.

Under the assumption of independent feature vectors, the log-likelihood of a model λ for a sequence of T audio feature vectors is computed as:

$$L(\lambda) = \frac{1}{T} \sum_{t=1}^T \log [p(\bar{\mathbf{x}}_t | \lambda)]. \quad (13)$$

In the present work we compute (13) for a sliding sequence of $T = 3$ vectors, as this allows the achieving of good time resolution in the identification of start and end timestamps for each *V. chilensis* vocalization.

Aiming at computational efficiency, we made use of a single background model to represent $p(\mathbf{X} | \lambda_{hyp})$. Using a GMM as the likelihood function, the background model λ_{UBM} is typically implemented as a large GMM trained to uniformly represent the distribution of audio features for the specific acoustic environment. Specifically, the background dataset was selected in such a way as to reflect the expected acoustic environment encountered during the operation of the *V. chilensis* recognizer.

The model λ_{hyp} for the target species S is obtained as adapted GMM, derived by adjusting the parameters of the background model λ_{UBM} via maximum *a posteriori* (MAP) estimation (Reynolds et al., 2000), using the training dataset for S . Since the dataset representing the target species is usually quite small, the MAP adaption only updates the mean vectors of the model. Because the target species model λ_{hyp} is created by updating the parameters in the background model λ_{UBM} , there is a close association between the corresponding mixture components of the two models. The last facilitates the discriminative capability of the species-specific recognizer.

The decorrelation stage as described by (5) permits the use of diagonal covariance mixture density GMMs that have fewer free variables in their covariance matrices to estimate and, therefore, are better trained to represent the target classes when limited training data are available (Reynolds et al., 2000).

Computing the log-likelihood ratio (LLR) (10) for a sequence of audio feature vectors \mathbf{X} requires estimation of the likelihood twice: for the target species model and for the background model. This is

repeated for each audio feature vector in \mathbf{X} and is computationally expensive for models with large number of mixtures. However the fact that the hypothesized species model λ_{hyp} was adapted from the background model λ_{UBM} allows the use of fast scoring methods (Reynolds et al., 2000; Saeidi, Sadegh Mohammadi, Ganchev, & Rodman, 2009). Since the components of the adapted GMM retain a correspondence with the mixtures of the background model, the acoustic feature vectors close to a particular mixture in the background model will also be close to the corresponding mixture in the target bird model. For that reason, instead of scoring all mixtures of the background model and the target model for each audio feature vector, we determine the top- C scoring mixtures in the background model and compute the background model likelihood using only these top- C mixtures (Reynolds et al., 2000). Next, we only score the audio feature vector against the corresponding $C = 10$ components in the adapted target species model to evaluate the likelihood. For a background model with M mixtures, this requires only $M + C$ computations per feature vector compared with $2M$ computations for the case of exhaustive likelihood ratio evaluation. In the present work we made use of the ALIZE implementation of the GMM-UBM (Benoit et al., 2007).

2.2.3. Score post-processing and decision making

Decisions computed via the likelihood ratio test are post-processed in order to estimate the sound event boundaries, that is the start and end timestamps of *V. chilensis* sound emissions, viz. isolated single-note call events and entire call series. Because in our study on the seasonal activity patterns of *V. chilensis* any split of call series into single-note calls or fragments composed of several calls is undesirable, we made use of a sequence of post-processing steps (Fig. 4).

The post-processing aims to merge together segments belonging to an isolated single-note call or to a call series emitted by single birds, pairs, or groups of *V. chilensis*. At each of these steps the algorithm operates on different time scales and applies certain domain knowledge about the structure of *V. chilensis* vocalizations (cf. Table 1). Such domain knowledge is related to the structure of *V. chilensis* vocalizations and the parameters minimum, maximum,

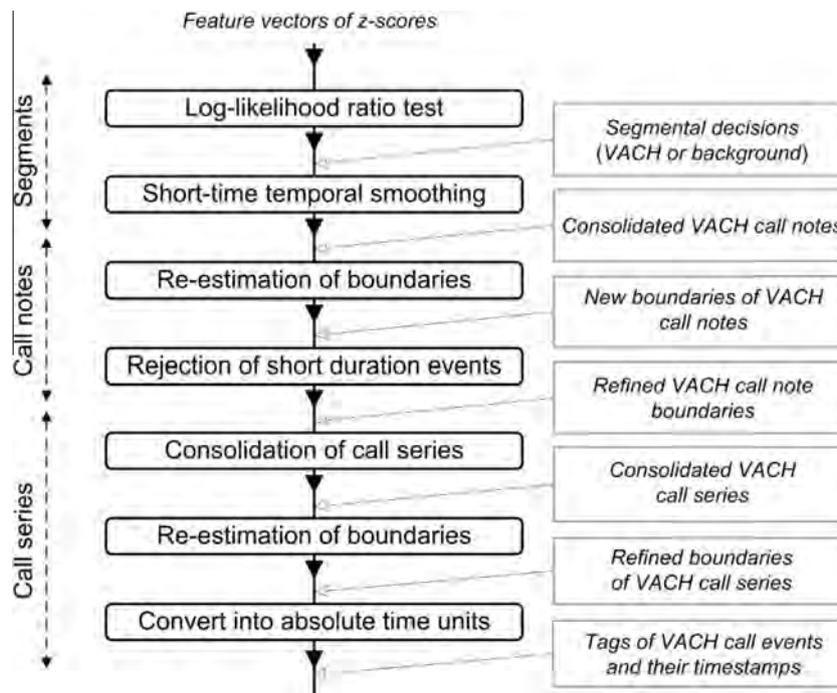


Fig. 4. Post-processing of the output of the log-likelihood ratio estimator.

Table 2

Domain knowledge used in the post-processing algorithm about the structure of isolated single-note call events and call series, by single birds, pairs, and groups of *Vanellus chilensis*.

	Duration [seconds]
<i>Isolated single-note calls</i>	
Minimum duration	0.10
Maximum duration	0.30
Estimated medium duration	0.15
Arbitrarily chosen minimum duration of pauses between distinct call events	>1.0
<i>Call series of a single bird</i>	
Minimum duration of pauses between call notes	0.05
Defined maximum duration of pauses between call notes, pertaining to the same call series	0.99
<i>Choruses</i>	
Minimum duration of overlapping call notes	0.10
Maximum duration of overlapping call notes	0.60
Estimated medium duration of overlapping call notes	0.30
Minimum duration of pauses between subsequent overlapping call notes	0.01
Defined maximum duration of pauses between subsequent overlapping call notes	0.99

and estimated medium duration of isolated call events as well as the duration of pauses between subsequent calls in call series of a single bird and in choruses (Table 2).

Admittedly, at this stage of development we made no effort to build a recognizer that can distinguish between call series/choruses emitted by single birds, pairs, or groups of *V. chilensis*. Likewise, overlapping vocalizations that, based on their largely differing amplitudes, were evidently emitted by several birds were sometimes merged within a single detection. This is because in the current study we make use of single-channel recordings, which do not facilitate the spatial localization and the acoustic separation of different birds.

In brief, each sequence of audio feature vectors \mathbf{X} , computed as explained in Section 2.2.1, is subject to the log-likelihood ratio test (10) applied on frame-by-frame basis. Subsequently, the scores obtained in this manner are compared with the decision threshold θ , resulting in a series of binary decisions: “1” – if the audio feature vector belongs to a *V. chilensis* call event or “0” – if the feature vector corresponds to sound of a different origin. Finally, these decisions are processed as follows in order to discover the boundaries and timestamps of the beginning and end of each call event.

For that purpose, we first convolve the frame-by-frame binary decisions with a rectangular window with a duration of 50 ms. This first step helps to avoid sporadic misclassifications of neighboring frames, which are due to short bursts of interfering audio signals. We keep for further processing only sequences with a length of at least 20 ms and thereby alleviate short events that are not likely to correspond to a *V. chilensis* vocalization.

In order to merge together closely spaced call events into a call series, we convolve the decisions made so far with a rectangular window of 1 s duration and estimate the call series boundaries.

Thus the sequence of post-processing steps constitutes the recognizer output which is converted to timestamps, indicating the onset and end of the *V. chilensis* call events. During the performance evaluation experiments the timestamps found by the recognizer were compared with the manually tagged ground-truth timestamps derived as explained in Section 2.1.3.

2.3. Experimental protocol and performance metrics

In all experiments we followed a common protocol that makes use of the datasets described in Section 2.1.3. Specifically, for the creation of the acoustic background model λ_{UBM} we made use of

approximately 27 h of audio recordings, representing the acoustic conditions specific to the area of the northern Pantanal. Next, using the method described in Section 2.2.2, we adapted a species-specific GMM model λ_{hyp} for *V. chilensis*. Collectively these two models were used for implementing the log-likelihood ratio test (10) for short sequences of three feature vectors. For estimating the value of the threshold θ (8) we made use of the four recordings, described as set VL01 (cf. Section 2.1.3), which contains a total of 80 call series of the target species. The decision threshold θ was set as the minimum value of θ that maximizes the recognition of *V. chilensis* call series with zero false positives on set VL01. Once the threshold θ was adjusted, the performance of the *V. chilensis* recognizer was evaluated with the validation dataset VL02, consisting of ten audio recordings with a total of 337 call series of the target species (cf. Section 2.1.3).

The recognition results were evaluated in terms of two performance metrics: *accuracy* (14) and *correct* (15) measured in percentages:

$$Accuracy = \frac{H - I}{N} \times 100, [\%], \quad (14)$$

$$Correct = \frac{H}{N} \times 100, [\%], \quad (15)$$

where H (*hits*) indicates how many times the *V. chilensis* recognizer correctly recognized a target vocalization, I stands for *insertions* (false positives), and N is the total number of target events according to the annotations of the test dataset VL02. The annotations of VL02 contain two types of *V. chilensis* events:

- 337 vocalizations confirmed by an expert ornithologist as originating from the target species *V. chilensis*,
- 137 vocalizations heard and/or seen in the spectrogram and allegedly originating from the target species but impossible to confirm acoustically with high confidence due to competitive sound events, high noise floor, or great distance of the calling bird from the microphone.

The events of the first type were considered as targets ($N = 337$) and were used when counting the number of hits (H) and misses ($N - H$) of the *V. chilensis* recognizer. The 137 events of the second group were merged with the 337 to form an extended set of $337 + 137 = 474$ *V. chilensis* events. This extended set was used as a reference when counting the number of insertions I . Therefore, for insertions we count only detections which do not coincide with the extended set of 474 *V. chilensis* events. However, due to the fact that many *V. chilensis* call series overlap with competing, sometimes very strong signals of other species, we inspected all recognizer-generated timestamps in Adobe Audition to exclude the possibility that an apparent hit in fact represented a false positive detection.

Since the validation dataset VL02 (Section 2.1.3) consists of unedited field audio recordings, the *V. chilensis* events have dissimilar amplitude and clarity. In the present study the main focus is on detecting call events emitted by birds within a range of few meters to several dozen meters from the microphone and with amplitude above -30 dB. These sound levels correspond to signal strengths that ornithologists can identify with certainty when using traditional audio visual survey methods (cf. Jahn, 2011a, 2011b) and thus facilitate the interpretation of the observed acoustic activity patterns.

However for the purpose of comprehensiveness, in Section 3 we report the recognition results at a wider range of sound levels: $[0, -20]$ dB, $[0, -30]$ dB, $[0, -40]$ dB, $[0, -50]$ dB. In fact, the category $[0, -50]$ dB also incorporates a certain number of call events with amplitude below -50 dB, which were discernible during manual annotation by an expert bioacoustician. These sound levels

Table 3

Split of the *Vanellus chilensis* vocalizations depending of the amplitude of the loudest note in a sound event.

Range	# Target	# Not confirmed	# Total
[0, –20] dB	44	0	44
[0, –30] dB	111	0	111
[0, –40] dB	240	2	242
[0, –50] dB	337	137	474

correspond to the maximum amplitude of the loudest note in each call event. The number of target events for each of these ranges is presented [Table 3](#).

In the reporting for each group, target events with amplitude below the specified lower boundary were not counted as misses if not detected by the *V. chilensis* recognizer.

3. Results

3.1. Detector performance assessment

All experimental results reported here are for the validation set VL02, which consists of ten 14-minute recordings (Section 2.1.3). In [Table 4](#) we present the recognition results for the *V. chilensis* recognizer for models of varying complexity: 1024, 512, 256, 128, 64, 32, 16 or 8 mixture components in the model, depending on the amplitude of the loudest call in each bird call series: [0, –20] dB, [0, –30] dB, [0, –40] dB, and [0, –50] dB.

The accuracy and correctly recognized target events depend on the size and the quality of the model ([Table 4](#)). All models achieved over 90% correct detections for *V. chilensis* vocalizations with high amplitude [0, –20] dB; however the models with 512 and 128 mixture components demonstrated superior accuracy as they allowed recognition of *V. chilensis* sounds with higher accuracy and without false positives (insertions) for all amplitude ranges. Although the accuracy results for the models with 512 and 128 mixture components are quite close, the performance of the model GMM128 is slightly better for signals with loudest calls in the range of primary interest [0, –30] dB as well as in the ranges [0, –40] dB and [0, –50] dB.

The results, namely recognition accuracy of 97.7% for strong *V. chilensis* sound events [0, –20] dB, and accuracy 83% for vocalizations with amplitude of the strongest call in the range [0, –30] dB, chart the margins of applicability of the current *V. chilensis* recognizer. Since the noise floor in natural environments is time-varying, many of the weaker calls and call series often get buried in the background noise. However we feel that this does not affect the usefulness of the recognizer since the acoustic activity of the lapwing is very well captured in the vicinity of the recording stations. In this context it is of particular importance that the reported accuracy is obtained on real-field recordings and for operational settings tuned at zero insertions, i.e., zero false positives rate. Avoidance of false positives is desirable as this makes the recognizer appropriate for automating biological studies focused on statistical analysis of presence/absence data and of hourly and daily *V. chilensis* vocal activity ([Figs. 5 and 6](#)).

3.2. Processing speed

On a contemporary PC with processor i7 working at 3.4 GHz and OS Microsoft Windows 7, the *V. chilensis* recognizer operates at 0.08 real time when a single core of the i7 processor is used. Further speed-up of computations is feasible by optimization of the audio pre-processing and audio feature extraction code, and by means of multi-core computing schemes. Such a speed-up will be the subject of further investigations.

Table 4

Recognition results depending on the loudest call amplitude in *Vanellus chilensis* vocalizations.

	[0, –20 dB]	[0, –30 dB]	[0, –40 dB]	[0, –50 dB]
GMM1024				
Hits, <i>H</i>	43	80	82	83
Misses, <i>N-H</i>	1	31	158	254
Insertions, <i>I</i>	0	0	0	0
Total instances, <i>N</i>	44	111	240	337
Accuracy	93.2%	85.6%	45.4%	32.9%
Correct	97.7%	87.4%	46.3%	33.5%
GMM512				
Hits, <i>H</i>	43	92	100	101
Misses, <i>N-H</i>	1	19	140	236
Insertions, <i>I</i>	0	0	0	0
Total instances, <i>N</i>	44	111	240	337
Accuracy	97.7%	82.9%	41.7%	30.0%
Correct	97.7%	82.9%	41.7%	30.0%
GMM256				
Hits, <i>H</i>	43	95	109	111
Misses, <i>N-H</i>	1	16	131	226
Insertions, <i>I</i>	3	3	3	3
Total instances, <i>N</i>	44	111	240	337
Accuracy	90.9%	82.9%	44.2%	32.0%
Correct	97.7%	86.5%	45.4%	32.9%
GMM128				
Hits, <i>H</i>	43	93	104	105
Misses, <i>N-H</i>	1	18	136	232
Insertions, <i>I</i>	0	0	0	0
Total instances, <i>N</i>	44	111	240	337
Accuracy	97.7%	83.8%	43.3%	31.2%
Correct	97.8%	83.8%	43.3%	31.2%
GMM64				
Hits, <i>H</i>	42	72	73	73
Misses, <i>N-H</i>	2	39	167	264
Insertions, <i>I</i>	0	0	0	0
Total instances, <i>N</i>	44	111	240	337
Accuracy	95.5%	64.9%	30.4%	21.7%
Correct	95.5%	64.9%	30.4%	21.7%
GMM32				
Hits, <i>H</i>	42	86	94	95
Misses, <i>N-H</i>	2	25	146	242
Insertions, <i>I</i>	1	1	1	1
Total instances, <i>N</i>	44	111	240	337
Accuracy	93.2%	76.6%	38.8%	27.9%
Correct	95.5%	77.5%	39.2%	28.2%
GMM16				
Hits, <i>H</i>	42	97	122	124
Misses, <i>N-H</i>	2	14	118	213
Insertions, <i>I</i>	6	6	6	6
Total instances, <i>N</i>	44	111	240	337
Accuracy	81.8%	82.0%	48.3%	35.0%
Correct	95.5%	87.4%	50.8%	36.8%
GMM8				
Hits, <i>H</i>	40	77	86	87
Misses, <i>N-H</i>	4	34	154	250
Insertions, <i>I</i>	8	8	8	8
Total instances, <i>N</i>	44	111	240	337
Accuracy	72.7%	62.2%	32.5%	23.4%
Correct	90.9%	69.4%	35.8%	25.8%

3.3. Analysis of detector results

The *V. chilensis* detections indicate a pronounced decrease in cumulated daily acoustic activity during July 2013 ([Fig. 6](#)). Furthermore, nocturnal activity is considerable higher during moonlight nights, i.e., between first quarter and last quarter moon compared with moonless nights around new moon. With respect to diurnal activity patterns there is a significant increase in the number of vocalization events near sunrise and sunset ([Fig. 5](#)), as

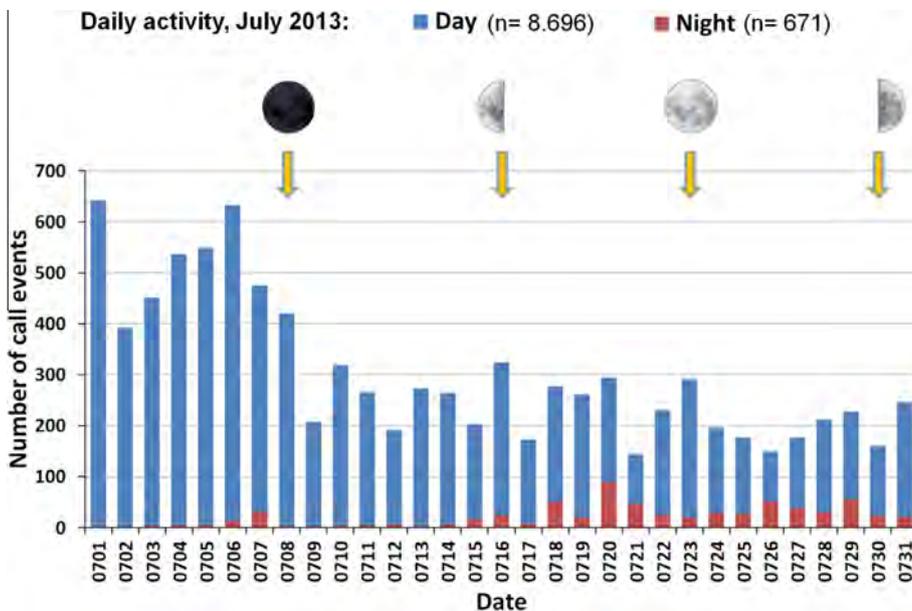


Fig. 5. Decrease in cumulated daily acoustic activity of *Vanellus chilensis* at the end of the breeding season in July 2013 and variation in nocturnal activity in relation of the lunar phases; recording station PPA001, Fazenda Pouso Alegre, Mato Grosso, Brazil.

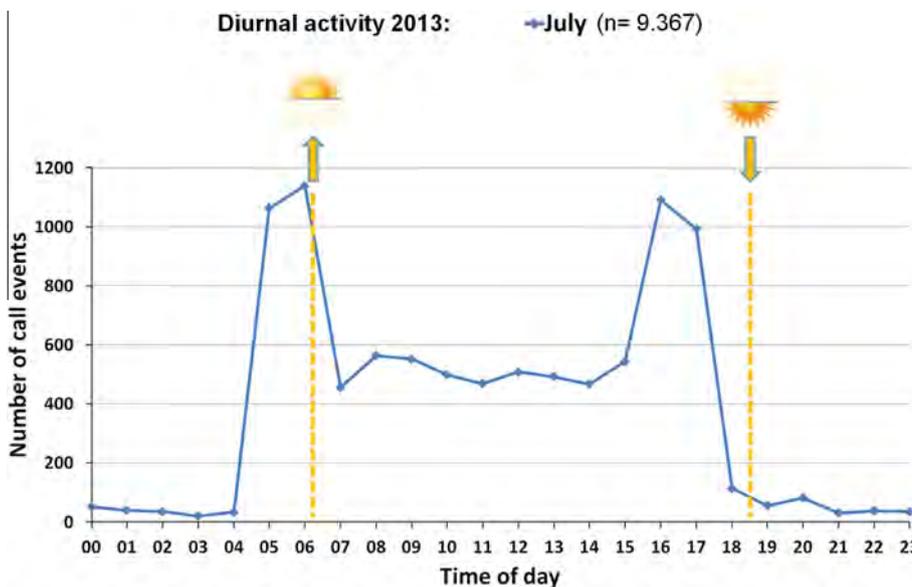


Fig. 6. Variation of cumulated diurnal acoustic activity of *Vanellus chilensis* in relation to sunrise and sunset in July 2013; recording station PPA001, Fazenda Pouso Alegre, Mato Grosso, Brazil.

is to be expected for bird species that are mostly diurnal (cf. Aide et al. (2013, p. 13, Fig. 4). The recognition results demonstrate that automated acoustic detection facilitates the quantitative assessment of changes in acoustic activity. In other words, detailed statistics on different time scales can be derived.

4. Discussion

4.1. Technology development aspects

The GMM-UBM approach employed here was originally developed for speaker verification tasks to determine whether or not a given sentence was pronounced by a specific person (Reynolds et al., 2000). The decision is usually made at the file level, based

on audio segments with durations in the range of a few seconds to some dozens of seconds or even minutes. The GMM-UBM approach was previously used for bird species classification only in a speaker verification-like setup (Graciarena, Delplanche, Shriberg, Stolcke, & Ferrer, 2010). Making use of audio recordings, assumed to contain a single bird species per file, Graciarena et al. (2010) tested each species-specific model with multiple true trials from the target species and with impostor⁶ trials from other species in order to compute the classification accuracy.

In contrast to Graciarena et al. (2010), our species-specific recognizer scans through long real-field recordings and detects start

⁶ Impostor (non-target) is when a species-specific model is tested against recordings of other species.

and end times of *V. chilensis* call events, without the need of preliminary segmentation of the audio to single-species segments. Furthermore, our method makes use of a purposely-developed post-processing algorithm, incorporating domain knowledge about the structure and duration of *V. chilensis* call events. This post-processing algorithm is applied on the output of the log-likelihood ratio estimator in order to obtain the boundaries and timestamps of single-note calls and call series.

In addition, while Graciarena et al. (2010) relied on GMM-UBM with 1024 components in order to study the applicability of different audio parameterization setups, we demonstrated that for the *V. chilensis* recognizer a smaller model with 128 components is more appropriate, even if we derived the *V. chilensis* model using an abundance of recordings for the target species (~45 min).

Kwan et al. (2006) created a complete system for localization, tracking, and classification of birds with multiple microphones. They reported 90% classification rates in experimental setups where bird sounds were simulated (played through a personal computer) and detected by their system. Later on, Marcarini, Williamson, and de Sisternes Garcia (2008) followed the same approach for GMM-based closed-set classification of nine bird species. The reported classification accuracy was in the range between 7.1% and 81.9% depending on the species and the acceptable level of false positive rates. Both works made use of plain GMM classification of birds, under the assumption that preliminary segmentation of the audio to single-species segments is available. The use of plain GMM classification does not allow for proper account of unlabeled co-occurring sound events and proper modeling of the acoustic background, which makes the classifier error prone in uncontrolled real-field conditions. By contrast, the GMM-UBM approach used in our work creates species-specific models, which are adapted from a large acoustic background model built for the specific operational environment. Therefore our modeling approach has the potential to achieve a better discriminative power in real-field conditions. Of course this advantage depends on the assumption that the acoustic background model is representative of the specific environment.

In the aforementioned studies (i.e., Graciarena et al. (2010), Kwan et al. (2006), Marcarini et al. (2008)) the experimental evaluation was performed with test files that contained only one dominant species per file, and the decision was made at file level without the need to estimate the boundaries of call events. In the present work, we evaluate the proposed method with continuous real-field recordings captured in uncontrolled conditions (cf. Sections 2 and 3) and automatically search for the call events of interest and their boundaries.

4.2. Practical use of the recognizer

The GMM-UBM based *V. chilensis* recognizer facilitates the analysis of acoustic activity patterns at different time scales. We demonstrate the practical usefulness of the recognizer with results on cumulated diurnal and daily acoustic activity of *V. chilensis* over one month of continuous recordings (Figs. 5 and 6). Comparable time resolution is hardly achievable with observer-based traditional bird survey methods, particularly in long-term monitoring studies.

In the near future we will screen complete annual cycles of recordings from multiple Pantanal recording stations with the *V. chilensis* recognizer. We hypothesize that the analysis of the distribution and abundance of the target species will show a close correlation between the annual flood-drought cycle in the Pantanal region and the life history of the lapwing. A research effort on such a scale is neither practical nor possible for human observers, as the collection of any comparable amount and quality of evidence is

logistically prohibitive, and most importantly, it is beyond the limits of traditional human-expert-based methods.

Autonomous acoustic recording units, in combination with automated species-specific recognizers, such as the one described here, facilitate the continuous 24/7 surveillance of sound-producing animal species and thereby provide valuable information on the biology and the conservation status of the species observed.

4.3. Research limitations

In the present study we aimed to achieve high recognition accuracy at zero false positives rates. Therefore we only considered *V. chilensis* vocalizations with loudest call notes in the range [0, -30] dB. Bird vocalizations at these sound levels can be undoubtedly confirmed by experienced ornithologists during traditional field surveys, even under noisy environmental conditions (Jahn, 2011a, 2011b). At the same time, a reliable automated recognition of weaker call events, with audio levels below -30 dB, would be of significant interest for biological studies, considering that human experts routinely detect animal vocalization down to -40 dB or even less (pers. obs.). However at the present development stage, any fine-tuning on the detection of such faint call events leads to a significant increase in the false positives rates. A major challenge to be addressed when targeting the detection of weak call events, with amplitude on or below the ambient noise floor, is that the recognition results heavily depend on the availability of reliable noise suppression methods and their efficiency in rapidly varying ambient noise conditions. At small distances of a few meters, multi-sensor microphone arrays provide the means for noise suppression, but at greater distances alternative approaches must be used.

A direct comparison of the automatically detected timestamps with the tags of the expert-annotated validation library VL02 revealed that in many cases only parts of longer call series were detected. In certain cases even some of the high-amplitude call notes went undetected. One reason for this phenomenon is the co-occurrence of sounds emitted by *V. chilensis* and by other species. The use of more advanced noise suppression methods, which operate off-line and are applied directly on the audio spectrogram seen as an image, may offer a solution to this problem (Briggs et al., 2012; Cadore, Gallardo-Antolan, & Pelaez-Moreno, 2011; Potamitis, 2014). However future research is required in order to investigate the applicability of these methods to the processing of long uncontrolled real-field recordings.

Finally, these acoustic models were built from a moderate amount of training (approx. 45 min) and background recordings (approx. 27 h), representing the acoustic conditions over a period of only a few months. In particular, for the background model we made use of fifty-four 14-minute recordings with a total duration of about 12 h from Fazenda Pouso Alegre soundscapes recorded in 2012, covering August (38 files), October (4), and November (12), as well as thirty-two 30-minute recordings (about 15 h) made in the SESC Pantanal Reserve area in January 2014. It is likely that this relatively small background library cannot represent the complex and variable Pantanal soundscapes of a complete annual cycle of recordings. Therefore, recognition accuracy over long time periods will depend on the availability of a representative acoustic background model, containing samples of all months of the year.

4.4. Follow-up technological development and research

Future research efforts will aim to increase the sensitivity of the *V. chilensis* recognizer in order to detect vocalizations down to amplitudes of at least -40 dB. The latter signal strengths are closer to the limits at which a bird-sound expert can identify

vocalizations under low-to-medium ambient noise conditions. Recent off-line noise suppression methods, which operate on the audio spectrogram seen as an image, offer promising options for the elimination of ambient noise (Briggs et al., 2012; Cadore, Gallardo-Antolan, & Pelaez-Moreno, 2011; Potamitis, 2014). These methods are usually based on morphological filtering of the spectrogram or make use of other established image noise reduction techniques in order to deal with short-duration sound events and/or long-duration interferences. Furthermore, embedding domain knowledge about the operational environment, the structure of the species-specific sound emissions, and other prior information are seen as promising research directions that could facilitate the detection of weaker call events. Likewise, rule- and grammar-based methods for post-processing the output of the log-likelihood ratio estimator could contribute to the improvement of the *V. chilensis* recognition results.

Other improvements, like the expansion of the acoustic background model, processing complete annual cycles of recordings, and the compilation of more comprehensive *V. chilensis* training libraries will be investigated as well. Specifically, we plan to investigate whether the enlargement and the enhancement of the *V. chilensis* training library with annotations at the level of call types and ethological functions could facilitate the analysis of the lapwings' behavior.

In order to implement long-term monitoring efforts, we foresee integration of the *V. chilensis* recognizer in the large Pantanal Database Repository created by the INAU Project 3.14.

5. Conclusion

The main advantage of the approach presented here is that we detect target species' call events and their timestamps in continuous real-field recordings, regardless of weather conditions and time of day. Related previous research focused on classification tasks at the file-level, that is, on the processing of short audio recordings containing sound signals of a single dominant species or on the processing of continuous soundscapes made in favorable climatic conditions, such as dry and calm weather.

The proposed automated method for the recognition of *Vanellus chilensis* call events incorporates three important components: (i) traditional audio parameterization, which is adapted to the frequency range of *V. chilensis* sound emissions; (ii) a statistical log-likelihood ratio estimator based on the GMM-UBM approach, which here is employed in a search-and-detect mode in order to find segments with acoustic activity of the target species; and (iii) post-processing of the log-likelihood estimator output in order to aggregate segments belonging to single-note calls and call series, and to estimate their boundaries in terms of absolute time. The recognizer demonstrated good sensitivity and ability to detect weak sounds with amplitudes down to -30 dB with low false positive rates and robustness to the variability of acoustic environments. Validation experiments on real-field recordings and analysis of recognizer results confirmed the practical usefulness of our approach. It provides the means to automatize ornithological studies and to perform statistical analysis of the acoustic activity patterns of *V. chilensis*.

We anticipate that the *V. chilensis* recognizer will be particularly useful in large-scale and long-term biological studies, as the automatically detected call events and their timestamps provide the basis for processing terabytes of recordings within reasonable time.

The automated detection of *V. chilensis* acoustic activity at zero false positive rates facilitates statistical tests at multiple time and geographical scales. For instance, it supports the detection of presence/absence patterns and the estimation of the target species' abundance in relation to the flood/drought cycle in our study area.

Future research will be focused on call-type level annotation of an expanded *V. chilensis* training library in order to support the analysis of the behavioral contexts in which certain vocalizations are emitted. This will make possible the determination of distinct phases of the species' natural history, namely start/end dates of territory establishment, copulations, nesting, presence of hatchlings and juveniles, and formation of flocks. Subsequently, specific behaviors could be correlated with other ecological factors such as climate and the water level data. It is clear that such a detailed knowledge would be very helpful for the implementation of meaningful management and conservation measures. Finally, an important research direction will be the increase of processing speed and data throughput of the recognizer. This will allow simultaneous processing of complete annual cycles of recordings recorded at multiple locations.

Beyond the horizon of our project, we expect that automated monitoring technologies will gain importance in global efforts to reduce biodiversity loss and avert dangerous degradation in the populations of sound-emitting animal species.

Author contributions

Collected and prepared audio recordings: MIM, KLS, OJ, and TG. Identified, analyzed, and tagged the data: OJ. Contributed reagents/materials/analysis tools: MIM, KLS, and JMF. Performed the experiments: TG. Validated the recognizer performance: TG and OJ. Wrote the paper: TG, OJ, MIM, JMF, and KLS.

Acknowledgments

The authors acknowledge with thanks the financial and logistic project support by the Brazilian Science Council (CNPq), the National Institute for Science and Technology in Wetlands (INAU/UFMT), the Brehm Foundation for International Bird Conservation, Germany, the project OP "Competitiveness" BG161PO003-1.2.04-0044-C0001 financed by the Structural Funds of the European Union, and the project ISP1 financed by the Technical University of Varna, Bulgaria. The authors acknowledge funds for laboratory and field equipment provided by Fundação de Amparo à Pesquisa do Estado de Mato Grosso (FAPEMAT), Brazil, Programa de Apoio a Núcleos de Excelência, (PRONEX) No. 009/2009.

The interns Dan Codea, Robert Koch, and Thibaut Bezaré kindly selected the recording segments (snippets) that served as the basis for the compilation of reference and training libraries. Luiz Vicente da Silva Campos Filho, Fazenda Pouso Alegre, and Director Waldir Wolfgang Valutky, Estância Ecológica SESC Pantanal: Reserva Particular do Patrimônio Natural, Baía das Pedras, generously allowed us to conduct our studies on their habitat-rich lands. Roseneide Soares kindly provided her immense administrative expertise during the project implementation phase. Last not least, special thanks go to the INAU directors Prof. Dr. Wolfgang Junk and Prof. Dr. Paulo Teixeira for their continuous support and encouragement throughout our field and laboratory work.

Appendix A

Source files for the generation of spectrograms. The start and end times refer to the 1.5-second segments shown in Fig. 1.

Spectrogram	Reference file	Start [m:ss.ms]	End [m:ss.ms]
a	BIAVCHCHVACHLA_BRMTAPPPA001C001A20120731T021500B_S001306E022415_ID1q1c1e0f0R	0:12.961	0:14.461
b	BIAVCHCHVACHLA_BRMTAPPPA001C001A20120727T054500B_S042800E050394_ID1q2c3e0f0R	0:16.353	0:17.853
c	BIAVCHCHVACHLA_BRMTAPPPA001C001A20120729T164500A_S142704E145800_ID1q3c3e0f0R	0:26.129	0:27.629
d	BIAVCHCHVACHLA_BRMTAPPPA001C001A20120731T174500A_S063200E080400_ID1q3c3e0f0R	0:14.788	0:16.288
e	BIAVCHCHVACHLA_BRMTAPPPA001C001A20120801T013000B_S031499E032000_ID1q2c2e0f0R	0:14.348	0:15.848
f	BIAVCHCHVACHLA_BRMTAPPPA001C001A20120730T023000B_S054014E064302_ID1q2c1e0f0R	0:15.883	0:17.383
g	BIAVCHCHVACHLA_BRMTAPPPA001C001A20120731T1101500A_S023000E032500_ID1q1c2e0f0R	0:26.352	0:27.852
h	BIAVCHCHVACHLA_BRMTAPPPA001C001A20120730T060000B_S110003E145800_ID1q1c3e0f0R	3:44.997	3:46.497

References

- Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., & Alvarez, R. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ*, 1, e103. <http://dx.doi.org/10.7717/peerj.103>.
- Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., & Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4), 206–214.
- Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.-H., & Frommolt, K.-H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, 31, 1524–1534.
- Brendle, A. (2003). Behind threats to world's largest freshwater wetland. national geographic news. January 10, 2003 (pp. 1–2). Retrieved March 10, 2015.
- Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S. J. K., et al. (2012). Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *Journal of the Acoustical Society of America*, 131, 4640–4650.
- Benoit, G., Fauve, B., Matrouf, D., Scheffer, N., Bonastre, J.-F., & Mason, J. S. D. (2007). State-of-the-art performance in text-independent speaker verification through open-source software. *IEEE Transactions on Audio, Speech and Language Processing*, 15, 1960–1968.
- Cadore, J., Gallardo-Antolan, A., & Pelaez-Moreno, C. (2011). Morphological processing of spectrograms for speech enhancement. In *Advances in nonlinear speech processing. Lecture notes in computer science* (Vol. 7015, pp. 224–231). Berlin Heidelberg: Springer.
- Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, 4, 675–683. <http://dx.doi.org/10.1111/2041-210X.12060>.
- Gill, F., & Donsker, D. (Eds.). (2014). IOC World bird list (v 4.1). <http://dx.doi.org/10.14344/IOC.ML.4.1>. URL: <www.worldbirdnames.org>, last accessed 7 February.
- Graciarena, M., Delplanche, M., Shriberg, E., Stolcke, A., & Ferrer, L. (2010). Acoustic front-end optimization for bird species recognition. In *Proceedings of ICASSP-2010* (pp. 293–296). Dallas, USA.
- Gwynne, J. A., Ridgely, R. S., Tudor, G., & Argel, M. (2010). *Birds of Brazil, Pantanal and Cerrado of Central Brazil* (Vol. 1). Ithaca, New York: Comstock Publishing Associates.
- Frommolt, K.-H., & Tauchert, K.-H. (2014). Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. *Ecological Informatics*, 21, 4–12.
- Internet Bird Collection. (2014). Southern Lapwing (*Vanellus chilensis*). In *The internet bird collection*. URL: <<http://ibc.lynxeds.com/species/southern-lapwing-vanellus-chilensis?t=1400664871>>, last accessed 12 May 2014.
- Jahn, O. (2011a). Surveying tropical bird communities: in search of an appropriate rapid assessment method. In K.-L. Schuchmann (Ed.), *Bird communities of the Ecuadorian Chocó: A case study in conservation* (pp. 63–107). Bonn, Germany: Bonner zoologische Monographien 56. Zoological Research Museum A. Koenig (ZFMK). URL: <http://zoologicalbulletin.de/BzB_Volumes/BzB_56/BzB_56.pdf>, last accessed 25 June 2014.
- Jahn, O. (2011b). Structure and organization of the bird community. In K.-L. Schuchmann (Ed.), *Bird communities of the Ecuadorian Chocó: A case study in conservation* (pp. 109–178). Bonn, Germany: Bonner zoologische Monographien 56. Zoological Research Museum A. Koenig (ZFMK). URL: <http://zoologicalbulletin.de/BzB_Volumes/BzB_56/BzB_56.pdf>, last accessed 25 June 2014.
- Junk, W. J., da Silva, C. J., da Cunha, C. N., & Wantzen, K. M. (2011). *The Pantanal: Ecology, biodiversity and sustainable management of a large Neotropical seasonal wetland*. Sofia, Bulgaria: Pensoft Publishers.
- Junk, W. J., & da Cunha, C. N. (2005). The Pantanal: A large South American wetland at a crossroads. *Ecological Engineering*, 24, 391–401.
- Härmä, A. (2003). Automatic identification of bird species based on sinusoidal modeling of syllables. In *Proceedings of the 2003 IEEE international conference on acoustics, speech, and signal processing, (ICASSP '03)* (Vol. 5, pp. 545–548).
- Henríquez, A., Alonso, J. B., Travieso, C. M., Rodríguez-Herrera, B., Bolaños, F., Alpizar, P., et al. (2014). An automatic acoustic bat identification system based on the audible spectrum. *Expert Systems with Applications*, 41, 5451–5465.
- Huang, C.-J., Yang, Y.-J., Yang, D.-X., & Chen, Y.-J. (2009). Frog classification using machine learning techniques. *Expert Systems with Applications*, 36, 3737–3743.
- Kwan, C. H., Ho, K. C., Mei, G., Li, Y., Ren, Z., Xu, R., et al. (2006). An automated acoustic system to monitor and classify birds. *EURASIP Journal on Advances in Signal Processing*, 1, 1687–1680. <http://dx.doi.org/10.1155/ASP/2006/96706>.
- Marchant, J., Prater, A. J., & Hayman, P. (1986). *Shorebirds – an identification guide to the waders of the world*. London, UK: Christopher Helm Publishers Ltd.
- Martin, J. P. (1997). The first Southern Lapwing *Vanellus chilensis* in Mexico. *Cotinga*, 8, 52–54.
- Maruyama, P. K., Cunha, A. F., Tizo-Pedroso, E., & Del-Claro, K. (2010). Relation of group size and daily activity patterns to Southern Lapwing (*Vanellus chilensis*) behavior. *Journal of Ethology*, 28, 339–344.
- Marcarini, M., Williamson, G. A., & de Sísternes Garcia, L. (2008). Comparison of methods for automated recognition of avian nocturnal flight calls. In *IEEE international conference on acoustics, speech and signal processing (ICASSP-2008)* (pp. 2029–2032). <http://dx.doi.org/10.1109/ICASSP.2008.451803>.
- Potamitis, I. (2014). Automatic classification of a taxon-rich community recorded in the wild. *PLoS ONE*, 9, e96936. <http://dx.doi.org/10.1371/journal.pone.0096936>.
- Potamitis, I., Ntalampiras, S., Jahn, O., & Riede, K. (2014). Automatic bird sound detection in long real-field recordings: Applications and tools. *Applied Acoustics*, 80, 1–9.
- Ridgely, R. S., & Gwynne, J. A. (1992). *A guide to the birds of Panama: with Costa Rica, Nicaragua, and Honduras*. NJ, USA: Princeton University Press.
- Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10(1–3), 19–41.
- Saeidi, R., Sadegh Mohammadi, H. R., Ganchev, T., & Rodman, R. D. (2009). Particle swarm optimization for sorted adapted Gaussian mixture models. *IEEE Transactions on Audio, Speech and Language Processing*, 17, 344–353.
- Santos, E. S. A. (2010). Southern Lapwing (*Vanellus chilensis*). In T. S. Schulenberg (Ed.), *Neotropical birds online. Ithaca: Cornell lab of ornithology*. URL: <http://neotropical.birds.cornell.edu/portal/species/overview?p_p_spp=144596>, last accessed 22 May 2014.
- Stowell, D. & Plumbley, M. D. (2010). Birdsong and C4DM: A survey of UK birdsong and machine recognition for music researchers. Technical report C4DM-TR-09-12. Centre for Digital Music, Queen Mary University of London.
- Stowell, D., & Plumbley, M. D. (2014). Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ*, 2, e488. <http://dx.doi.org/10.7717/peerj.488>.
- Sueur, J., Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PLoS ONE*, 3, e4065. <http://dx.doi.org/10.1371/journal.pone.0004065>.
- Schuchmann, K.-L., Marques, M. I., Jahn, O., Ganchev, T., & de Figueiredo, J. M. (2014). Os sons do Pantanal: um projeto de monitoramento acústico automatizado da biodiversidade. *O Biólogo, Revista do Conselho Regional de Biologia*, 29, 12–15.
- Stotz, D. F., Fitzpatrick, J. W., Parker, T. A., III, & Moskovits, D. K. (1996). *Neotropical birds: Ecology and conservation*. Chicago, USA and London, UK.
- Wetlands International, 2010. State of the World's Waterbirds, 2010. (Compiled by Simon Delany, Szabolcs Nagy, and Nick Davidson). Wetlands International, Ede, The Netherlands. Retrieved March 2014. URL: <http://www.wetlands.org/Portals/0/publications/Report/SOWW2010%20%283%29.pdf>.
- Wiersma, P. (1996). Southern Lapwing (*Vanellus chilensis*). In J. del Hoyo, A. Elliott, & J. Sargatal (Eds.), *Handbook of the birds of the world. Hoatzin to Auks* (Vol. 3). Barcelona, Spain: Lynx Editions, p. 420.